DOCUMENT RESUME

ED 053 185 TM 000 715

AUTHOR	Pennell, Roger
TITLE	Finding Points of View in Judgment Data.
INSTITUTION	Educational Testing Service, Princeton, N.J.
REPORT NO	RB-71-8
PUB DATE	Feb 71
NOTE	18p.
EDRS PRICE	EDRS Price MF-$0.65 HC-$3.29
DESCRIPTORS	Algorithms, *Behavioral Science Research, *Computer Oriented Programs, *Data Analysis, *Individual Differences, *Mathematical Models, Mathematics

ABSTRACT

It is argued that many investigators utilize the Tucker and Messick (1963) Model with no intention of looking for individual differences or, after utilizing the model, draw improper inferences. An example is given illustrating the difficulties which result from improper use of the model. Several proper methods are outlined. (Author)

RB-71-8

RESEARCH BULLETIN

FINDING POINTS OF VIEW IN JUDGMENT DATA

Roger Pennell

This Bulletin is a draft for interoffice circulation. Corrections and suggestions for revision are solicited. The Bulletin should not be cited as a reference without the specific permission of the author. It is automatically superseded upon formal publication of the material.

Educational Testing Service
Princeton, New Jersey
February 1971

FINDING POINTS OF VIEW IN JUDGMENT DATA

Roger Pennell

Educational Testing Service

Abstract

It is argued that many investigators utilize the Tucker and Messick (1963) Model with no intention of looking for individual differences or, after utilizing the model, draw improper inferences. An example is given illustrating the difficulties which result from improper use of the model. Several proper methods are outlined.

FINDING POINTS OF VIEW IN JUDGMENT DATA

Roger Pennell[1]

Educational Testing Service

I. Introduction

It certainly must be argued that the availability of computers to experimenters in the behavioral sciences provides the capability for much finer and much more thorough data analyses. With the myriad of multivariate procedures which are more or less routinely implemented on our computers an investigator finds himself confronted with a large number of tacks he might take to evaluate his experimental hypothesis. Often, however, the investigator shortchanges himself by utilizing the most exotic of procedures. The case in point is the model by Tucker and Messick (1963), henceforth TM, to analyze a data matrix of p judgments by N subjects into components accounting for subject variance and components accounting for judgment variance. Whereas before, one could only wonder about individual differences that were known to exist in a sample of subjects, one now had a procedure to isolate the components of these individual differences. Whereas before, one analyzed the mean judgment (or every subject's set of judgments separately) the sample could be partitioned into groups giving more or less homogeneous responses.

The thesis propounded in this paper is, first, that investigators tend to have misconceptions concerning the model, and second (not necessarily as a result of the first), investigators tend to misuse the model. In order to operate on common ground let us digress and indicate the exact model.

The most common utilization of the TM model occurs when an investigator has obtained $n(n - 1)/2 = p$ judgments on all pairs of n stimuli for each of N subjects. This generates the $p \times N$ data matrix X assumed to have the following form:

$$X = UGW \tag{1}$$

where U contains the column-wise eigenvectors of XX' , W contains the row-wise eigenvectors of $X'X$, and G is a diagonal matrix containing the positive square roots of the eigenvalues of either XX' or $X'X$.

Due to a theorem by Eckart and Young (1936) we know that for any arbitrary rank r we necessarily produce a least squares approximation to X by

$$X_r = U_r G_r W_r \quad , \tag{2}$$

where X_r is least squares, rank-r approximation to X , U_r contains the first r columns of U , W_r the first r rows of W and G_r the first r rows and columns of G . The experimenter usually chooses r by one or another subjective procedure aimed at finding the minimum "significant" number of components needed in the model. At this point Tucker and Messick state that the elements in U_r represent projections of stimulus pairs on unit length principal vectors of X , the elements of W represent projections of people on the unit length principal vectors of X and that, further, each column of U represents a set of distance measures for the set of p judgments. We can now, for instance, absorb G_r into U_r and W_r and produce a transformation on W_r , say T , that is more psychologically pleasing than the principal vector orientation and still preserve the form of the model as

(3) $$X_r = (U_r G_r^{\frac{1}{2}} T^{-1})(T G_r^{\frac{1}{2}} W_r) = YZ \quad .$$

Perhaps the most interesting notion that TM develop is that of an idealized individual. Since the columns of Z represent projections of people on r rotated dimensions, it is clear that we may append any number of additional columns (representing imagined or idealized individuals), say m of them, on to the end of Z and after premultiplying by Y, our matrix X_r will be P by $N + m$ where the last m columns represent judgments made by idealized individuals. As such these judgments may be analyzed by one or another multidimensional scaling routine to obtain the underlying structure of the stimuli as they appear to the idealized individuals.

We shall proceed in three phases: to show two common misuses of the model; to use a set of artificial data to show that incorrect interpretations are a result of these misuses; and to illustrate the proper approach to analyzing such data.

II. Misuses of the Model

TM state that one should expect the first component of U to be highly correlated with the mean judgment, which brings us to our first point. Knowing that the first component of U essentially represents a set of mean judgments, some investigators apply the TM point-of-view routine with no intention of searching for individual differences in their data. With some phrase like "the pattern of eigenroots was inspected and it was decided that one component was sufficient to ...," they could analyze the distance from only the first component and simultaneously report the utilization of a fancy multivariate

procedure. It is argued that this procedure is wrong for three reasons: (a) The rationale for selecting only one component is usually related to the very large size of the first eigenroot. It was clearly stated in the TM paper that we should expect the first eigenroot to be large (due to choosing not to eliminate means variance by row-centering) and that this state of affairs is totally independent of whether or not individual differences exist. (b) Only in the most uninteresting of cases (certainly null) is it tenable to assert that there exist no consistent, identifiable characteristics of subjects which produce intersubject variance. (c) Granted that we have rightly or wrongly decided to eliminate considerations of individual differences, why use the elements of an eigenvector to represent distance measures when we can put our feet on the ground with actual means with known sampling properties?

The second area of conceptual difficulty centers around the notion that the decomposition in (2) provides us with individual points of view, or individual sets of distance measures which can each be analyzed to obtain representative stimuli configurations. No matter whether one considers U_r or Y, the column-wise elements are not in general all positive and therefore do not even possess the elementary property of distances: nonnegativeness. Some would argue that a set of distances both positive and negative simply constitutes an "additive constant" problem; however, this author has had little interpretive success upon scaling such numbers based on this premise.

A helpful heuristic in conceptualizing the subject space is to consider it made up of a large number of directions. As we move along some particular direction some facet of stimulus relationships changes in a consistent fashion. As an example, subjects closer to the origin in a particular direction might

perceive stimulus i and j to be closer together than subjects farther from the origin in this same direction. Were we to pick a point in the space, multiply through its coordinates to get an idealized set of distances and find that some of these distances were negative, we should be satisfied that we have chosen an idealized subject that we could never, even theoretically, observe. This is so because he perceives two or more stimuli as being so close together that their distance is negative. It seems at best fatuous to analyze distances from a subject who is theoretically not observable. Furthermore, taking, say, the i -th column of U_r as a set of distance measures is equivalent to utilizing the one-dimensional centroid (mean) of the corresponding i -th subject component from W_r . That is to say, this is one way of idealizing the i -th component of subject variance. But, indeed, this is the height of absurdity unless there exist subjects with high scores on the i -th component of W_r and essentially zero scores on all other components. If this is not the case, we are implicitly embracing a model which says that the way in which subjects make judgments about stimuli can be viewed as a multidimensional process, and that we are interested in one dimension of that process even though it produces judgments not at all like the judgments actually made. For this reason the statement made by TM: "These stimulus-pair projections, when ... rotated to orientations possibly more appropriate psychologically than the principal-axes position, will constitute measures of distance between pairs of stimuli" (Tucker & Messick, 1963, p. 29), is simply not worded strongly enough, i.e., we must isolate dimensions, by means of rotation, which pass through clusters of real subjects, and, as such, generate an essentially "simple structure" space for subjects.[2] Without this we embrace the somewhat bizarre model alluded to above. We shall delay this point until the example

which follows and acknowledge that Cliff (1968) has cogently argued a rather similar point.

III. Example

As an example we shall consider a fictitious set of data in which rather extreme points of view actually exist. We shall generate points of view by concocting four ways in which a set of two-dimensional stimuli might be "conceptualized" by hypothetical subjects. Figure 1a represents a standard conceptualization, 1b and 1c represent subjects that use either the first dimension or the second, but not both, 1d represents a uniform contraction of the 1a space. This example is slightly extreme, but it is not hard to imagine a population of subjects that differ in their perceptions of a set of stimuli along the lines of Figure 1. The four sets of interpoint distances corresponding to the four points of view about the stimuli were computed, and an additional sample of four subjects was generated for each of the points of view by adding random noise distributed as $N(0.5)$ to each "true" interpoint distance. This generates the matrix X as $p = 28$ and $N = 20$ (five subjects for each point of view). X was decomposed by (1) and (2) taking $r = 4$. The elements of G were 1070.79, 455.34, 10.57, 7.12, 6.61, 4.21, 3.28, 2.89, 2.66, 2.49, 2.28, 1.90, 1.78, 1.57, 1.33, 1.17, .87, .77, .67, .56. If these roots were derived from exploratory data, one would surely not take more than three components; on the other hand, one should not conclude that there is only one point of view merely because there is one enormously large root. Presumably there appear to be only three points of view because the first and last population points of view are so similar.

Insert Figure 1 about here

What happens if we decide to use the elements of the first eigenvector of U_r as measures of the interpoint distances of the eight points? We can get a feeling for what kind of configuration we are going to obtain by considering the correlation of this vector with the four sets of true interpoint distances obtained from Figure 1. The correlations in order are .9757, .8852, .3325, and .9367; the multiple correlation between the four sets of distances and the first vector is .9999. It seems clear that the set of interpoint distances we are considering scaling (the first eigenvector of U_r) is exactly a linear combination of the distances we should be concerned with (the true distances) but is imperfectly correlated with any one of them, i.e., the first eigenvector of U_r is a figment of our imagination and represents no empirical state of affairs whatsoever.

Results such as obtained from our first eigenvector of U_r make evident the folly of the "normative" approach to research in the behavioral sciences. Indeed, what good is it to "predict and control" behavior of a normed non-existent entity? Clearly we can discard the "first eigenvector" approach to resolving the data matrix.

What of the second, third and fourth eigenvectors of U_r , is there any hope of finding a correspondence with the original set of distances? Table 1 presents a rectangular correlation matrix where rows represent the last three eigenvectors of U_r and columns represent the four sets of interpoint distances from Figure 1. Here it looks like the second vector is a bipolar representation of the second and third viewpoints; however, the other viewpoints are not evident. In any case we should expect a virtually unconditional identification since we started from concocted data, and the results in Table 1 do not afford

such identification. In no case can we hope to recover configurations of stimuli like those in Figure 1, even though we know them to be present, from the last three vectors of U_r .

Insert Table 1 about here

IV. Admissible Sets of Distances

In our case neither rules of thumb nor orthogonal rotations will yield an admissible set of distances--a set which correlates almost perfectly with the original set, and which, therefore, affords the possibility of recovering the exact configurations of stimuli. We have to simply look at the data (W_r) and observe that there are four clusters of points (subjects) lying on obliquely related axes. The problem can be attacked in either of two ways. We can, as Cliff (1968) suggests, merely read off the centroids of those four clusters, array each centroid as a column in, say, D , and produce

(4) $$X_r^* = U_r G_r D \; ,$$

where X_r^* represents judgments of distance made by four idealized individuals. Note that this essentially averaging process (in computing the centroids) is not subject to the same philosophical criticism as using a mean vector to represent the judgments of all the subjects Here we have presumably isolated the components of individual differences, and, as well, groups of subjects that consistently respond alike. We can therefore argue that using centroids is a very natural way to deal with the measurement error that we expect.

A second way to attack the problem, based on our knowledge of which subjects belong to which groups, is to produce a pattern matrix, P, representing group membership. In this case the matrix would be $N \times 4$ and the i-th row would contain a 1 in the column representing the group to which the i-th subject belongs and a 0 in all other columns. The matrix to use in (4), D^* say, is then found to be

$$D^* = PW_r'(W_rW_r')^{-1}W_r \quad , \tag{5}$$

or

$$D^* = PW_r'W_r \tag{5a}$$

since $W_rW_r' = I$. Note that $PW_r' = T$ from (3) and that one obtains the distances corresponding to the groups from the matrix Y.

Using this approach on our artificial data the distances in the columns of Y have correlations of .994, .992, 1.00 and .972 with the respective original distances. Clearly, if our scaling algorithm is sufficiently precise, we can be confident of retrieving the input configurations.

The method utilizing the D^* matrix is possibly the most versatile in practice. If the number of groups is large we need not go to the trouble to plot subject points and gauge the extent to which they cluster, rather we need only gauge the extent of agreement between P and D^*. The extent to which they agree reflects the extent to which we have been able to find a nonrigid rotation of the subject axes such that they pass through clusters of actual subjects. Here we would be willing to tolerate small negative values in Y as long as the fit between D^* and P was quite good.

It should be noted that using this approach we are strictly unable to locate a number of groups, say g, which is less than r. This is the case because what we need is the left-hand inverse of T, which doesn't exist when g is less than r. In our example the four groups fell out rather nicely because they were the four salient components of subject variance and therefore came out as mixtures of the first four principal components. The initially appealing idea of taking r to be large, perhaps the full set of components, and trying to find, say, two components representing male judgments and female judgments is, for the above reason, doomed to fail. If we take only two components, $r = 2$, and thus ensure g not less than r, we are most unlikely to have these two components represent any mixture of sex variance whatsoever, i.e., it would be extremely unlikely that sex differences would be prominent enough to come out as the first two components unless the experimental task was explicitly designed to contrast sex differences.

It should be pointed out that the rather typical problem in these types of analyses, especially when the sample of subjects is large, is that when trying to plot the subject points in r-dimensional space we find one large, irregularly shaped cluster of points. Using the rationale developed to this point one clearly proceeds along one of two lines: Decide that the individual differences are uninteresting or at least unsystematic and therefore compute mean judgments and scale those, or take the judgments of these subjects who seem to span the cluster of subject points and scale each in turn. One thereby determines how internalized representations of the stimuli vary as the range of individual differences contained in the sample is spanned.

V. Summary

We have tried to argue that simplistic and/or heuristic approaches to the TM model are often inadequate. In particular, there is apparently little to recommend the utilization of the first eigenvector as a set of distance judgments.

References

Cliff, N. The "idealized individual" interpretation of individual differences in multidimensional scaling. Psychometrika, 1968, 33, 225-232.

Eckart, C., & Young, G. The approximation of one matrix by another of lower rank. Psychometrika, 1936, 1, 211-218.

Tucker, L. R, & Messick, S. An individual differences model for multidimensional scaling. Psychometrika, 1963, 28, 333-367.

Footnotes

[1]The author is indebted to Robert Weber, Cornell University, for performing the necessary computer programing.

[2]This is not to say that one may not eliminate the rotation problem altogether by choosing interesting points corresponding to idealized individuals.

Table 1

Correlations between Eigenvectors of

U_r and Original Distances, D

	D_1	D_2	D_3	D_4
U_2	-.2215	-.7288	.7826	-.2793
U_3	-.0467	-.0913	-.0961	.2096
U_4	.1167	.1827	.1874	.3205

Figure Caption

Fig. 1. Four hypothetical "conceptualizations" about 8 stimuli in 2-space.

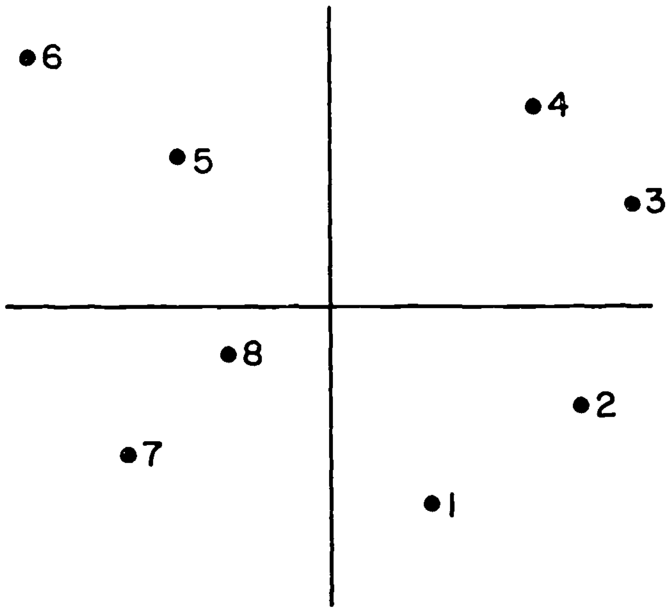

Fig. 1a

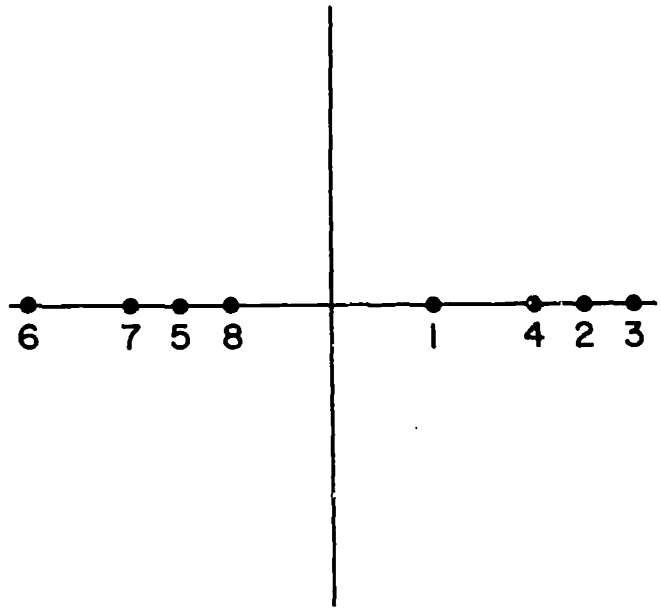

Fig. 1b

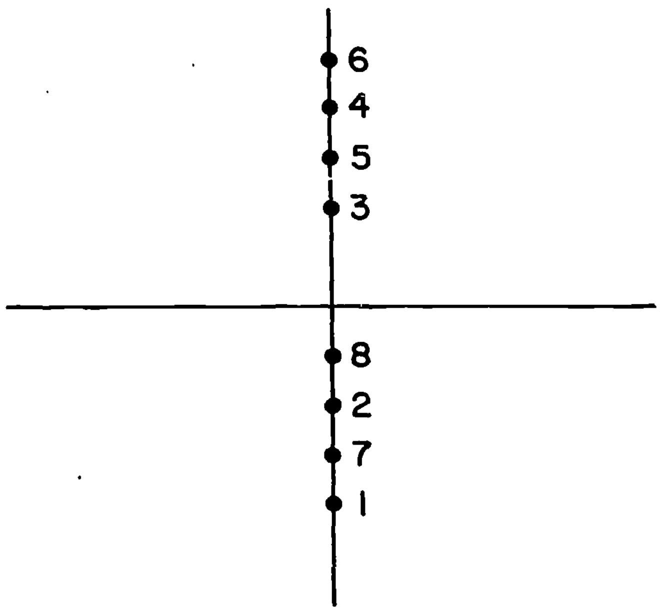

Fig. 1c

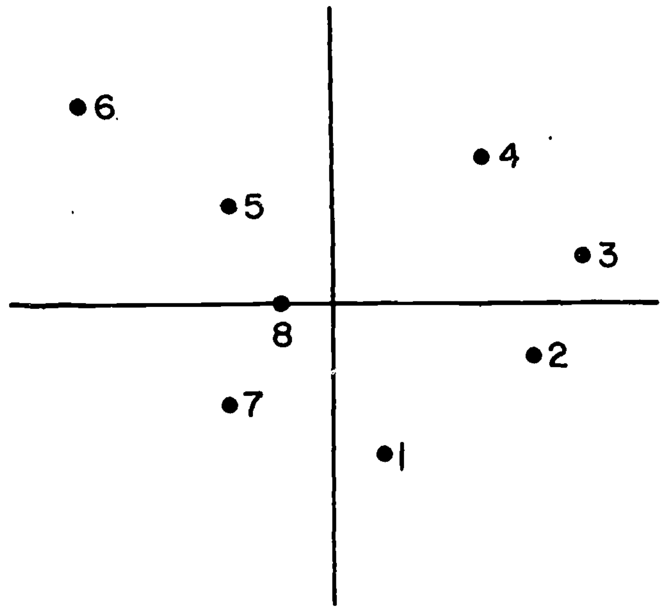

Fig. 1d